

Econometría

Capítulo II: Regresión Lineal

Índice de contenido

1. Regresión lineal simple matemática	1
1.1. Introducción al modelo matemático	1
1.2. Bondad del ajuste	2
1.3. Regresiones no lineales reducibles a lineales.....	3
2. Regresión lineal simple estadística	4
2.1. Introducción al modelo estadístico	4
2.2. Estimación.....	6
2.3. Inferencia	6
3. Regresión lineal múltiple estadística.....	8
3.1. Estimación por mínimos cuadrados.....	8
3.2. Coeficiente de determinación	9
3.3. Modelo de Gauss-Markov en regresión múltiple	10
3.4. Quitar y añadir variables.....	15
3.5. Inferencia	18
Bibliografía.....	22

*Pese a que este documento podría ser suficiente material de estudio, siempre
recomiendo que se consultes el material oficial.*

*Si encuentras alguna errata o error, por favor házmelo saber por medio de mi correo
electrónico daniel.garcia.garcia@alumnos.upm.es o mi Telegram @dgarcia_97.*

1. Regresión lineal simple matemática

1.1. Introducción al modelo matemático

El objetivo será, dado un conjunto de puntos conocidos (x_i, y_i) en el plano \mathbb{R}^2 , obtener la ecuación de una recta (*porque es una regresión lineal*) que aproxime mejor a dicha nube de puntos. La recta que se desea obtener tiene la forma:

$$y = a + bx$$

A la variable x se le llama **variable explicativa**, regresora o exógena, mientras que a la variable y se le denomina **variable explicada**, dependiente o endógena. Al parámetro a se llama intercepto y al b pendiente de la recta de regresión.

Como se ha indicado al inicio, la recta busca aproximar los valores de la nube de puntos, por lo que en general no pasa exactamente por los puntos de la nube. Por esta razón se genera lo que se denomina residuos, que básicamente son la diferencia entre el valor de la nube de puntos y el valor de la recta que trata de aproximarlos.

$$e_i = \overset{\text{punto}}{\hat{y}_i} - \overset{\text{aproximación de la recta}}{(a + bx_i)} = y_i - a - bx_i$$

Para obtener la recta utilizamos el método de mínimos cuadrados, que consiste en minimizar la suma de los cuadrados de los residuos (*se eleva al cuadrado para dejar en positivo los errores negativos*). Para ello se define la función $g(a, b)$ sobre la que se trabajará para hallar su valor mínimo:

$$g(a, b) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - a - bx_i)^2$$

La solución al problema depende de qué variable se haya escogido como explicada y explicativa. La solución para x como explicativa e y como explicada no es la misma que en el caso inverso.

El mínimo de la función $g(a, b)$ se obtiene cuando su derivada es cero (*pendiente nula*):

$$\begin{cases} \frac{\partial g}{\partial a} = -2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \\ \frac{\partial g}{\partial b} = -2 \sum_{i=1}^n x_i (y_i - a - bx_i) = 0 \end{cases}$$

Despejando se obtiene:

$$\begin{cases} an + b \sum x_i = \sum y_i \\ a \sum x_i + b \sum x_i^2 = \sum (x_i \cdot y_i) \end{cases}$$

Estas ecuaciones se denominan ecuaciones normales, y su solución es la siguiente:

$$a = \frac{\bar{y} \sum x_i^2 - \bar{x} \sum x_i y_i}{\sum x_i^2 - n \bar{x}^2}$$

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

Donde \bar{x} e \bar{y} son las medias muestrales de x e y . Una alternativa para obtener a de una manera más sencilla es calcularla en base a b gracias a que sabemos que la recta pasa por el punto (\bar{x}, \bar{y}) :

$$a = \bar{y} - b\bar{x}$$

El sistema es compatible si hay variación muestral en x , es decir, que no todos los valores x_i sean iguales. Si todos los valores de x_i son iguales, entonces la recta de regresión se reduce a un punto si asignamos x como la variable explicativa, pues sólo toma un punto. Siempre que la variable explicativa tenga variación muestral, se confirma que el punto (a, b) candidato a mínimo es efectivamente un mínimo. Ahora basta con sustituir sus valores en la ecuación $y = a + bx$ para obtener la denominada **recta de regresión**.

Como se sabe que la recta pasa por el punto (\bar{x}, \bar{y}) , podemos escribir la recta de regresión como:

$$y - \bar{y} = b(x - \bar{x})$$

1.2. Bondad del ajuste

Para determinar la calidad de la aproximación la recta de regresión a la nube de puntos estudiamos el **coeficiente de determinación R^2** , que es una medida adimensional que mide cuan mala es la aproximación obtenida.

$$R^2 = \frac{SSE}{SST} = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2}$$

Se cumple que $0 \leq R^2 \leq 1$. Cuanto más se aproxima a 1, más bondad del ajuste de la recta de regresión, mientras que cuanto más se aproxime a 0 peor es. El método de mínimos cuadrados calcula la recta con mayor R^2 posible. Además, también podemos interpretar R^2 como el porcentaje de la variación muestral de y que es explicada por x . Para hallar dicho porcentaje basta con multiplicar R^2 por cien:

$$\%_{\text{explicación}} = R^2 \cdot 100$$

Se denomina **coeficiente de correlación lineal de Pearson R** a la raíz cuadrada de R^2 , aunque con un matiz: R siempre debe tener el signo de la pendiente de la recta de regresión pese a que la raíz cuadrada de R^2 arroje un valor positivo.

1.3. Regresiones no lineales reducibles a lineales

Existen casos en los que la relación entre la variable explicada y la variable explicativa no es lineal. En algunos de esos casos podemos utilizar técnicas que nos permitan convertir el problema no lineal en lineal, de manera que podamos hacer uso del método de mínimos cuadrados para resolverlo. Dos ejemplos frecuentes de esta situación son:

- **Regresión logarítmica:** La relación parece seguir curva logarítmica, por lo que se escoge la ecuación $y = a + b \ln x$. En este caso podemos aplicar el método de mínimos cuadrados simplemente definiendo a la muestra como $(\ln x_i, y_i)$.
- **Regresión exponencial:** La relación para seguir una curva exponencial, por lo que se escoge la ecuación $y = ce^{bx}$, $c > 0$. Para convertirlo en lineal aplicamos logaritmos, de tal forma que la ecuación de regresión es $\ln y = \ln c + bx$. A continuación, aplicamos el método de mínimos cuadrados sobre la muestra definida como $(x_i, \ln y_i)$.

2. Regresión lineal simple estadística

2.1. Introducción al modelo estadístico

En el modelo matemático los datos de la muestra que se obtienen son absolutos y deterministas, es decir, por muchas veces que observemos los datos éstos siempre serán los mismos. Puede ser útil emplear el modelo matemático en ocasiones donde se conozca con certeza todos los datos. Sin embargo, en la mayoría de los casos del mundo real no se disponen de todos los datos de forma certera.

Ejemplos:

No se disponen de todos los datos cuando tratamos de modelizar la cantidad de cosechas (y) en base a la cantidad de abono empleado (x). Si usamos el modelo matemático o determinista, empleando una cantidad x_i de abono en sucesivos experimentos debería resultar en la misma cantidad y_i en cada experimento. Sin embargo, existen otras variables desconocidas o imprevisibles como el clima. Una tormenta podría reducir el valor de y_i , arrojando un valor diferente dado el mismo x_i . Se hace entonces necesario un modelo que contemple el hecho de que hay variables que no pueden predecirse y, por tanto, conocerse.

Otro ejemplo pueden ser los resultados de ventas. Los datos sobre los que trabajamos dependen de las tiendas y períodos seleccionados. Por tanto, se hace necesario un modelo que contemple el hecho de que estamos simulando el total de las ventas con una muestra que no representa a la población.

Entonces distinguimos entre un **modelo verdadero** o poblacional (*determinista*) que es inaccesible, donde son conocidos todos los datos certeramente en la nube de puntos, y un **modelo muestral** (*no determinista*) donde sólo se conoce una cantidad limitada e incompleta de datos del modelo verdadero, pero a través de los cuales se intenta simular este. Para ello debemos considerar las propiedades de la población como variables aleatorias, es decir, como datos que a priori se desconoce su valor.

Los datos del modelo matemático anteriormente visto no tienen una ecuación que los genere sencillamente porque todos son conocidos. Sin embargo, en el modelo muestral los datos se generan a través de una ecuación que denominamos **ecuación del modelo**. Esta ecuación es

$$y = \beta_1 + \beta_2 x + \varepsilon,$$

donde y es la variable explicada, x la variable explicativa, β_1 el intercepto, β_2 el coeficiente de x y ε el residuo que representa todos aquellos factores desconocidos del modelo verdadero, siendo así el responsable de las variaciones de y dado el mismo valor de x .

x , y y ε son variables aleatorias porque no conocemos todos sus valores del modelo verdadero. β_1 y β_2 tampoco son conocidos, pero se sabe que tienen el mismo valor en todo el modelo verdadero, por lo que no se van a considerar variables aleatorias, sino parámetros que deseamos estimar a través de la muestra recogida.

Los datos del modelo muestral se generan dando valores a la variable x en la ecuación del modelo y observando los resultados del resto de variables aleatorias. En primer lugar, se determina el número n de experimentos u observaciones que se realizarán. Para obtener cada resultado de y y ε definimos los valores de x del modelo muestral, lo que significa que antes de realizar generar la muestra conocemos todos los valores de x que vamos a dar, aunque se siga considerando variable aleatoria, pues las x del modelo verdadero que no se toman en el muestral son desconocidas. Dar valores a x en el modelo muestral implica que cualquier otra muestra que se genere debe usar esos mismos valores.

De esta manera la muestra queda definida por el siguiente sistema de ecuaciones:

$$\begin{cases} y_1 = \beta_1 + \beta_2 x_1 + \varepsilon_1 \\ y_2 = \beta_1 + \beta_2 x_2 + \varepsilon_2 \\ \dots \\ y_n = \beta_1 + \beta_2 x_n + \varepsilon_n \end{cases}$$

Simplificando la notación del sistema de ecuaciones obtenemos

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \quad i = 1, \dots, n.$$

El modelo que se va a estudiar en esta se denomina **modelo de Gauss-Markov**. Este modelo impone las siguientes condiciones o hipótesis:

1. **Linealidad:** Los datos se generan mediante una muestra definida por $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, $i = 1, \dots, n$, donde β_1 y β_2 son parámetros reales constantes desconocidos y x_i valores constantes conocidos que tendrán los mismos valores en cualquier otra muestra. Recordemos que las variaciones en las muestras son a causa de la variable aleatoria ε_i . El tamaño de la muestra n está fijado.
2. **Variación muestral de la variable explicativa:** Aunque sean conocidos, no todos los valores de x_i son iguales. Esto significa que puede haber valores repetidos de x pero no todos. Se cumple, por tanto, que $\sum (x_i - \bar{x})^2 > 0$.
3. **Perturbación con media cero:** Las perturbaciones $\varepsilon_1, \dots, \varepsilon_n$ son variables aleatorias con media cero, es decir, $E[\varepsilon_i] = 0$, $i = 1, \dots, n$.
4. **Homocedasticidad:** Las varianzas de las perturbaciones $\varepsilon_1, \dots, \varepsilon_n$ existen y es son todas iguales a σ^2 , es decir, $E[\varepsilon_i^2] = \sigma^2$, $i = 1, \dots, n$.
5. **No correlación:** Todos los posibles pares de perturbaciones $(\varepsilon_i, \varepsilon_j)$ están incorrelados, es decir, $E[\varepsilon_i \varepsilon_j] = 0$, $i, j = 1, \dots, n$, $i \neq j$.
6. **Distribución de los errores:** Las perturbaciones se generan a través de una distribución normal con esperanza cero, es decir, $\varepsilon_i \sim NID(0, \sigma)$. Esto significa que el modelo muestral depende del valor de σ , que no es conocido, pero puede estimarse.

Llegamos a este punto queda definido el entorno de trabajo. El objetivo de la regresión lineal estadística es calcular una estimación de los tres parámetros β_1 , β_2 y σ sobre los que depende el modelo muestra a través de los estimadores $\hat{\beta}_1$, $\hat{\beta}_2$ y s .

2.2. Estimación

El estimador $\hat{\theta}$ de un parámetro θ de una variable aleatoria y es un estadístico obtenido a partir de una muestra y_i que permite realizar una estimación de θ . Para calcular los estimadores del modelo muestral (*suponiendo que es un modelo Gauss-Markov*) usamos el método de los mínimos cuadrados. Obtenemos que

$$\hat{\beta}_1 = \frac{\bar{y} \sum (x_i^2 - \bar{x}) \sum (x_i y_i)}{\sum (x_i^2 - n \bar{x}^2)},$$

$$\hat{\beta}_2 = \frac{\sum ((x_i - \bar{x})(y_i - \bar{y}))}{\sum ((x_i - \bar{x})^2)}.$$

Los estimadores son variables aleatorias porque depende de la muestra, que como se ha visto se representa con el sistema de ecuaciones $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, $i = 1, \dots, n$.

Cuando se calcula un estimador, se busca tres propiedades deseables:

1. **Insesgadez:** Un estimador $\hat{\theta}$ de la variable aleatoria θ es insesgado si su esperanza coincide con el valor del parámetro a estimar, es decir, $E[\hat{\theta}] = \theta$.
2. **Eficiencia:** La eficiencia de un estimador se mide por su varianza, lo ideal para un estimador insesgado es que sea lo más pequeña posible.
3. **Consistencia:** Un estimador es consistente si a medida que aumenta el tamaño de la muestra n el estimador tiende al valor del parámetro a estimar: $\lim_{n \rightarrow \infty} E[\hat{\theta}] = \theta$ y $\lim_{n \rightarrow \infty} V[\hat{\theta}] = 0$.

Es conveniente que un estimador sea insesgado, y entre los insesgados, sea lo más eficiente posible.

El **teorema de Gauss-Markov** para regresión simple afirma que los estimadores $\hat{\beta}_1$ y $\hat{\beta}_2$ son estimadores lineales insesgados de máxima eficiencia, es decir, con menor posible varianza entre todos los estimadores lineales de β_1 y β_2 . Un estimador lineal insesgado de máxima eficiencia se denomina por sus siglas en inglés BLUE (*Best Lineal Unbiased Estimator*).

Su demostración se realizará en regresión lineal múltiple estadística, al ser ésta general.

2.3. Inferencia

La inferencia se estudiará con más detalle en regresión lineal múltiple estadística al ser más general.

Un modelo de regresión trata de explicar la variación de la variable dependiente en términos de variaciones de la variable independiente. Esto tiene sentido únicamente cuando y está relacionada con x , es decir, cuando $\beta_2 \neq 0$.

Esta sección consiste en aplicar un test sobre la hipótesis nula $H_0: \beta_2 = 0$ en contraste con la alternativa $\beta_2 \neq 0$. La hipótesis nula se rechaza si $\hat{\beta}_2$ difiere significativamente de cero. Dado que $\hat{\beta}_2$ proviene de una variable aleatoria es importante considerar que existe cierto grado de incertidumbre. Para ello escalamos $\hat{\beta}_2$ en base a su desviación estándar σ . Además, para poder aplicar el test sobre la hipótesis nula es necesario conocer la distribución de $\hat{\beta}_2$.

El test que aplicaremos en este caso será el test t-Student. Definimos como el error estándar de $\hat{\beta}_2$ a

$$s_{\hat{\beta}_2} = \frac{s}{\sqrt{\sum (x_i - \bar{x})^2}},$$

donde $s = \sqrt{\frac{1}{n-2} \sum \hat{\varepsilon}_i^2}$ es el error estándar de la regresión y $\hat{\varepsilon}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_i$.

El t-valor de β_2 se define como

$$t_{\hat{\beta}_2} = \frac{\hat{\beta}_2}{s_{\hat{\beta}_2}}.$$

Además, s^2 es un estimador insesgado de máxima eficiencia de la varianza de la perturbación σ^2 . Demostraremos esta afirmación en regresión lineal múltiple estadística.

3. Regresión lineal múltiple estadística

3.1. Estimación por mínimos cuadrados

Partimos directamente del modelo estadístico surgido de tomar las muestras con k variables explicativas x_2, \dots, x_k ($x_1 = 1$ porque su coeficiente es el intercepto). Recordemos que las variables x son conocidas, pues se establecen como los valores para realizar los experimentos antes de ejecutarlos. La ecuación que lo define es

$$y = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon.$$

Cuando se realizan n experimentos del modelo estadístico obtenemos una muestra de este construyendo así un sistema de ecuaciones tal que

$$\begin{cases} y_1 = \beta_1 + \beta_2 x_{21} + \dots + \beta_k x_{k1} + \varepsilon_1 \\ y_2 = \beta_1 + \beta_2 x_{22} + \dots + \beta_k x_{k2} + \varepsilon_2 \\ \dots \\ y_n = \beta_1 + \beta_2 x_{2n} + \dots + \beta_k x_{kn} + \varepsilon_n \end{cases},$$

expresado de otro modo como

$$y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i, \quad i = 1, \dots, n.$$

Recordemos que las variables y_i y ε_i ya no representan variables aleatorias, sino resultandos del experimento i . Este sistema de ecuaciones de los experimentos del modelo puede reescribirse matricialmente¹ como

$$y = X\beta + \varepsilon,$$

donde

$$y = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{21} & \dots & x_{k1} \\ 1 & x_{22} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{2n} & \dots & x_{kn} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{pmatrix}, \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}.$$

Las filas de las matrices representan datos del experimento i , donde $i = 1, \dots, n$. El vector de residuos ε se puede definir de la siguiente manera despejando la ecuación:

$$\varepsilon = y - X\beta.$$

¹ Cuidado con confundir la variable y del modelo con la matriz de misma notación y que representa y_i .

Los experimentos empíricos sólo muestran los datos de la nube de puntos (X, y) , por lo que los coeficientes de β sólo pueden estimarse en base a dicha nube. El resultado es obtener un vector de estimaciones $\hat{\beta}$. Puede observarse que ε dependen de β , y como su valor sólo puede estimarse por el estimador $\hat{\beta}$, entonces los valores ε de también serán estimaciones. Como resultado obtenemos el estimador $\hat{\varepsilon}$.

$$\hat{\varepsilon} = y - X\hat{\beta}.$$

El objetivo será calcular un hiperplano que aproxime a valores de y en base a los datos extraídos del experimento. Para esto, primero es necesario hallar el estimador $\hat{\beta}$ a través del método de mínimos cuadrados.

Como deseamos obtener un estimador $\hat{\beta}$ insesgado cumpliendo con el modelo de Gauss-Markov, se impone que la suma de los cuadrados de los residuos sea mínima. Se define la función

$$g(\hat{\beta}) = \sum \hat{\varepsilon}_i^2 = \hat{\varepsilon}'\hat{\varepsilon} = (y - X\hat{\beta})'(y - X\hat{\beta}) = \dots = yy' - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}.$$

Se deriva la función de mínimos cuadrados con respecto al vector $\hat{\beta}$ y se iguala a cero para minimizar $g(\hat{\beta})$, obteniendo así las **ecuaciones normales**:

$$\left(\frac{\partial g}{\partial \hat{\beta}_i}\right) = -2X'y + 2X'X\hat{\beta} = 0 \Rightarrow X'X\hat{\beta} = X'y.$$

Basta despejar para obtener la ecuación que permite calcular $\hat{\beta}$:

$$\hat{\beta} = (X'X)^{-1}X'y.$$

Para que $\hat{\beta}$ exista, la matriz $X'X$ debe tener invertible o regular y por tanto X también debe serlo, es decir, $Rg(X) = k$. Para que esto se cumpla es necesario que $n \geq k$, es decir, el número de experimentos debe ser mayor al número de parámetros (*demostración algebraica, pero no se hará*). Una vez obtenido el vector estimador $\hat{\beta}$, obtenemos el hiperplano de regresión empleando la expresión

$$y = \hat{\beta}_1 + \hat{\beta}_2x_2 + \dots + \hat{\beta}_kx_k.$$

3.2. Coeficiente de determinación

El **coeficiente de determinación R^2** permite determinar cuán bueno es el ajuste de la muestra al hiperplano de regresión. Se define como

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sum (y_i - \bar{y})^2}.$$

El coeficiente de determinación debe estar comprendido entre 0 y 1, matemáticamente $0 \leq R^2 \leq 1$. Cuanto más cercano a 1 más bueno es el ajuste de la muestra al hiperplano. Sin embargo, al aumentar el número de variables explicativas, aumenta artificialmente el coeficiente de determinación, por lo que por convenio haremos uso del **coeficiente de determinación ajustado \bar{R}^2** que previene este incremento. Se define como

$$\bar{R}^2 = 1 - \frac{n-1}{n-k}(1 - R^2).$$

3.3. Modelo de Gauss-Markov en regresión múltiple

Para analizar las propiedades estadísticas de la estimación por mínimos cuadrados se emplea, de la misma manera que en regresión lineal simple estadística, el modelo de Gauss-Markov, pero ampliado a múltiples variables explicativas. Las hipótesis son:

1. **Linealidad:** Los datos sobre los que se trabaja son generados mediante una muestra empírica de n experimentos del modelo poblacional. Esta muestra queda expresada como $y = X\beta + \varepsilon$.
2. **No colinealidad:** La matriz X de constantes conocidas tiene rango k . Esto implica que la matriz $X'X$ es regular y por tanto invertible. Como se ha visto antes esto hace necesario que $n \geq k$.
3. **Perturbación con media cero:** La media de los errores o perturbaciones es cero, es decir, $E[\varepsilon] = E[\varepsilon_i] = 0$.
4. **Homocedasticidad:** La matriz de covarianza de los errores $E[\varepsilon\varepsilon']$ existe y todos los elementos de su diagonal son iguales a σ^2 en toda la muestra, es decir, $E[\varepsilon\varepsilon'] = E[\varepsilon_i^2] = \sigma^2$.
5. **No correlación:** Las perturbaciones correspondientes a diferentes experimentos no están correlacionadas entre sí. Por tanto, los elementos ajenos a la diagonal de la matriz de covarianza de errores $E[\varepsilon\varepsilon']$ son cero, es decir, $E[\varepsilon_i\varepsilon_j] = 0$ donde $i \neq j$ e $i, j = 1, \dots, n$.
6. **Distribución de los errores:** La perturbación es una distribución normal independientemente distribuida con esperanza cero, es decir, $\varepsilon_i \sim NID(0, \sigma)$. Se verifica a través de la hipótesis 5.

Combinando las hipótesis 4 y 5 se deduce que la matriz de covarianza de errores $\text{var}[\varepsilon] = E[\varepsilon\varepsilon'] = \sigma^2 I$, donde I es la matriz identidad.

Observamos que el modelo depende de $k + 1$ parámetros: β y σ . A continuación, calcularemos el estimador insesgado más eficiente de cada uno junto con el de la perturbación ε .

❖ Teorema de Gauss-Markov y demostración.

Para la regresión múltiple, el teorema de Gauss-Markov define que $\hat{\beta}$ es el vector de estimadores lineales insesgados de máxima eficiencia, es decir, con menor posible varianza. A continuación, se demostrarán las tres afirmaciones.

1. Demostración de linealidad. Un estimador γ es lineal si es una combinación lineal de y , de forma que puede expresarse como $\gamma = Ay$ siendo A una matriz $k \times n$ no aleatoria. Como deseamos demostrar que el vector de estimadores $\hat{\beta}$ es lineal, tomamos primero su expresión,

$$\hat{\beta} = \overbrace{(X'X)^{-1}X'}^A y.$$

Como puede observarse, podemos expresarlo como $\hat{\beta} = Ay$ donde $A = (X'X)^{-1}X'$. Dado que la matriz A es una matriz compuesta por operaciones entre los valores de las variables explicativas, es no aleatoria y de tamaño $k \times n$. Queda demostrado que $\hat{\beta}$ es lineal.

2. Demostración de insesgader. Recordemos que un estimador es insesgado si se cumple que la esperanza del estimador es igual al valor que se pretende estimar. Es decir, $E[\hat{\theta}] = \theta$. Si $\hat{\beta}$ es insesgado, significa que $E[\hat{\beta}] = \beta$.

Aplicando las hipótesis del modelo de Gauss-Markov sobre la ecuación de $\hat{\beta}$ y sustituyendo y por su expresión en el modelo completo obtenemos que

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \varepsilon) = (X'X)^{-1}(X'X\beta + X'\varepsilon) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon = \beta + (X'X)^{-1}X'\varepsilon,\end{aligned}$$

De este resultado se sabe que son variables aleatorias β y ε . Aplicaremos sobre ambas la siguiente propiedad: Sea una expresión lineal de la forma $z = A_1z_1 + A_2z_2$, donde z_1 y z_2 son variables aleatorias y A_1 y A_2 dos variables no aleatorias, las cuatro variables de las dimensiones correctas de manera que la expresión quede bien definida, tiene la siguiente propiedad: $E[z] = A_1E[z_1] + A_2E[z_2]$.

Por tanto, considerando $z = \hat{\beta}$, $A_1 = 1$, $z_1 = \beta$, $A_2 = (X'X)^{-1}X'$ y $z_2 = \varepsilon$, se cumple la condición

$$E[\hat{\beta}] = E\left[\overbrace{1}^{A_1} \overbrace{\hat{\beta}}^{z_1} + \overbrace{(X'X)^{-1}X'}^{A_2} \overbrace{\varepsilon}^{z_2}\right] = E[\beta] + ((X'X)^{-1}X')E[\varepsilon].$$

Se sabe que $E[\beta] = \beta$ y por la hipótesis 3 que $E[\varepsilon] = 0$, por tanto,

$$E[\hat{\beta}] = \beta + ((X'X)^{-1}X') \cdot 0 = \beta,$$

quedando de esta manera demostrado que $E[\hat{\beta}] = \beta$ y como consecuencia el estimador $\hat{\beta}$ es insesgado.

3. *Demostración de máxima eficiencia.* Que un estimador sea de máxima eficiencia significa que el resto de posibles estimadores tendrán una varianza mayor, y por tanto mayor incertidumbre. Para demostrar que $\hat{\beta}$ es el estimador de máxima eficiencia, partimos de su expresión como combinación lineal y sustituimos y por su expresión en el modelo completo. Obtenemos que

$$\hat{\beta} = Ay = A(X\beta + \varepsilon) = AX\beta + A\varepsilon.$$

donde $A = (X'X)^{-1}X'$. La varianza de un estimador lineal insesgado γ ,

$$\text{var}(\gamma) = \text{var}(Ay) = \text{var}(AX\beta + A\varepsilon),$$

sólo depende de la varianza de ε dado que es el único vector aleatorio; $AX\beta$ es una matriz de números reales, no aleatorios, por lo que su varianza es cero. Por lo tanto, obtenemos que

$$\text{var}(\gamma) = \text{var}(A\varepsilon) = E[(A\varepsilon - E[A\varepsilon])(A\varepsilon - E[A\varepsilon])].$$

Como $E[A\varphi] = AE[\varphi]$ donde φ es un vector aleatorio, $E[\varepsilon] = 0$ y $E[\varepsilon\varepsilon'] = \sigma^2$ por las hipótesis de Gauss-Markov, obtenemos que

$$\begin{aligned} E[(A\varepsilon - E[A\varepsilon])(A\varepsilon - E[A\varepsilon])] &= E[(A\varepsilon - AE[\varepsilon])(A\varepsilon - AE[\varepsilon])] \\ &= E[(A\varepsilon)(A\varepsilon)] = E[A'A\varepsilon'\varepsilon] = A'AE[\varepsilon'\varepsilon] = \sigma^2 A'A. \end{aligned}$$

En definitiva,

$$\text{var}(\gamma) = \sigma^2 A'A.$$

A continuación, comparamos la variación de cualquier estimador insesgado γ con el estimador insesgado de supuesta máxima eficiencia $\hat{\beta}$. Recordando que $\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$ obtenemos que

$$\text{var}(\gamma) - \text{var}(\hat{\beta}) = \sigma^2 A'A - \sigma^2(X'X)^{-1} = \sigma^2(A'A - (X'X)^{-1}).$$

Si se afirma que $\text{var}(\gamma) - \text{var}(\hat{\beta})$ es una matriz semidefinida positiva, entonces significa que $A'A - (X'X)^{-1} \geq 0$, o con otras palabras, que $A'A - (X'X)^{-1} = B$ sea también una matriz semidefinida positiva, que lo es (*no se demostrará*).

Al estimador lineal insesgado más eficiente se le denomina BLUE por sus siglas en inglés (*Best Lineal Unbiased Estimator*).

❖ **Estimación del estimador insesgado más eficiente: $\hat{\varepsilon}$.**

El modelo estimado se representa por la expresión

$$y = X\hat{\beta} + \hat{\varepsilon} \Rightarrow \hat{\varepsilon} = y - X\hat{\beta}.$$

Sobre esta ecuación podemos sustituir $\hat{\beta}$ por su expresión de manera que queda

$$\hat{\varepsilon} = y - X(X'X)^{-1}X'y = My,$$

donde $M = I - X(X'X)^{-1}X'$. La matriz M tiene una serie de propiedades características: es simétrica ($M' = M$) e idempotente ($M^2 = M$). Además, $MX = 0$, ya que

$$(I - X(X'X)^{-1}X')X = X - X(X'X)^{-1}X'X = X - X = 0.$$

M es una matriz que se utilizará posteriormente con frecuencia para acortar los cálculos y por sus propiedades anteriormente descritas.

Sustituyendo y por su expresión del modelo completo obtenemos

$$\hat{\varepsilon} = My = M(X\beta + \varepsilon).$$

Como $MX = 0$ obtenemos que

$$\hat{\varepsilon} = M\varepsilon.$$

Aplicando esperanzas y sabiendo por las hipótesis del modelo de Gauss-Markov que $E[\varepsilon] = 0$ obtenemos

$$E[\hat{\varepsilon}] = E[\varepsilon] - X(X'X)^{-1}X'E[\varepsilon] = 0,$$

determinando de esta manera que $\hat{\varepsilon}$ es un estimador insesgado de ε , ya que sus esperanzas son iguales: $E[\varepsilon] = E[\hat{\varepsilon}] = 0$.

❖ **Estimación del estimador insesgado más eficiente: $\hat{\beta}$.**

Aplicando el teorema de Gauss-Markov podemos afirmar que el estimador $\hat{\beta}$ obtenido por mínimos cuadrados a través de la ecuación

$$\hat{\beta} = (X'X)^{-1}X'y.$$

es el estimador insesgado de máxima eficiencia de β .

❖ **Estimación del estimador insesgado más eficiente: σ .**

A continuación, calcularemos el estimador s del vector aleatorio desconocido restante, la varianza de la perturbación σ^2 . Para ello debemos empezar con el cálculo de la varianza del estimador de la perturbación $\hat{\varepsilon}$, obtenemos así

$$\text{var}(\hat{\varepsilon}, \hat{\varepsilon}) = E[(\hat{\varepsilon} - E[\hat{\varepsilon}])(\hat{\varepsilon} - E[\hat{\varepsilon}])].$$

Sabemos que $E[\hat{\varepsilon}] = 0$ por lo que sustituyendo queda que

$$\text{var}(\hat{\varepsilon}, \hat{\varepsilon}) = E[\hat{\varepsilon}'\hat{\varepsilon}] = \text{tr}(E[\hat{\varepsilon}'\hat{\varepsilon}]) = \sigma^2 \text{tr}M,$$

donde $M = I - X(X'X)^{-1}X'$. Sabiendo que las trazas cumplen la propiedad de que $\text{tr}(A + B) = \text{tr}A + \text{tr}B$ y que $\text{tr}(A - B) = \text{tr}(B - A)$, obtenemos que

$$\begin{aligned} \text{tr}M &= \text{tr}(I_n - X(X'X)^{-1}X') = \text{tr}(I_n) - \text{tr}(X(X'X)^{-1}X') = n - \text{tr}(X'X(X'X)^{-1}) \\ &= n - \text{tr}(I_k) = n - k. \end{aligned}$$

Sustituyendo el resultado obtenemos que

$$\text{var}(\hat{\varepsilon}, \hat{\varepsilon}) = E[\hat{\varepsilon}'\hat{\varepsilon}] = \sigma^2(n - k).$$

De esta ecuación poder extraer un estimador insesgado s^2 de σ^2 tal que

$$s^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n - k}.$$

De s^2 obtenemos s de manera que

$$s = \sqrt{\frac{\hat{\varepsilon}'\hat{\varepsilon}}{n - k}}.$$

Este estimador s se denomina error estándar de la regresión. Además, con él podemos completar el cálculo de la matriz de covarianzas de $\hat{\beta}$, $\text{var}(\hat{\beta}, \hat{\beta})$, al sustituir σ^2 por su estimación insesgada s^2 . Recordemos que $\text{var}(\hat{\beta}, \hat{\beta}) = \sigma^2(X'X)^{-1}$, por lo que sustituyendo queda

$$\text{var}(\hat{\beta}, \hat{\beta}) = s^2(X'X)^{-1}.$$

Aplicando cada coeficiente $\hat{\beta}_j$ de $\hat{\beta}$ podemos obtener el denominado error estándar del coeficiente $\hat{\beta}_j$, que viene expresado por su matriz de covarianzas. Obtenemos que

$$\text{var}(\hat{\beta}_j, \hat{\beta}_j) = s^2 a_{jj},$$

donde a_{jj} es el elemento jj de la matriz $(X'X)^{-1}$.

3.4. Quitar y añadir variables

A continuación, se estudiará el efecto que tiene añadir o quitar variables del modelo descrito por el sistema de ecuaciones en forma matricial $y = X\beta + \varepsilon$, donde X es la matriz de variables explicativas con $Rg(X) = k$.

El modelo anterior puede expresarse como un modelo cuyas variables explicativas están divididas dos grupos, uno con $k - g$ variables explicativas (*variables que no se desean eliminar*) y otra las g variables restantes (*las que se desean eliminar o añadir*). Quedaría como

$$y = X\beta + \varepsilon \Rightarrow y = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

También puede expresarse considerando $\beta = (\beta_1 \ \beta_2)$ y $X = (X_1 \ X_2)$, obteniendo

$$y = (X_1 \ X_2) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \varepsilon.$$

Este modelo se denomina completo mientras que el resultante de eliminar las variables explicativas contenidas en X_2 (y por tanto $\beta = (\beta_1)$) se denomina modelo restringido.

❖ Estimación en el modelo restringido.

En el modelo restringido se calcula el estimador de β_1 , único coeficiente que hay, al que denotaremos como $\hat{\beta}_R$. Como en este modelo $\beta = (\beta_1)$, sólo se emplea la matriz de coeficientes X_1 y no X , por lo que

$$\hat{\beta}_R = (X_1'X_1)^{-1}X_1'y.$$

Para calcular el estimador del error obtenemos

$$\hat{\varepsilon}_R = y - X_1\hat{\beta}_R.$$

El modelo estimado resultante queda como

$$y = X_1\hat{\beta}_R + \hat{\varepsilon}_R.$$

❖ Estimación en el modelo completo.

En este modelo se utilizan las ecuaciones que hasta ahora hemos visto para hallar estimadores. Para el estimador de β obtenemos

$$\hat{\beta} = (X'X)^{-1}X'y.$$

Para el estimador del error obtenemos

$$\hat{\varepsilon} = y - X\hat{\beta}.$$

❖ Comparación entre: $\hat{\beta}_1$ y $\hat{\beta}_R$.

El modelo completo como el restringido tienen en común un determinado número de variables explicativas y sus respectivos coeficientes. Los coeficientes correspondientes a estas variables en común vienen dados por el vector β_1 , ya que $\beta = (\beta_1 \ \beta_2)$ para el completo y $\beta = (\beta_1)$ para el restringido.

A continuación, se estudiará la diferencia entre los estimadores de los coeficientes del conjunto común de ambos modelos β_1 , es decir, la diferencia entre $\hat{\beta}_1$ (del modelo completo) y $\hat{\beta}_R$ (del modelo restringido). Para ello se expresa $\hat{\beta}_R$ en relación con $\hat{\beta}_1$. Esto se logra sustituyendo en la ecuación de $\hat{\beta}_R$ la variable y por su ecuación del modelo completo, esto es

$$\begin{aligned}\hat{\beta}_R &= (X_1'X_1)^{-1}X_1'y = (X_1'X_1)^{-1}X_1'\overbrace{(X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\varepsilon})}^y \\ &= ((X_1'X_1)^{-1}X_1'X_1\hat{\beta}_1) + ((X_1'X_1)^{-1}X_1'X_2\hat{\beta}_2) + ((X_1'X_1)^{-1}X_1'\hat{\varepsilon}).\end{aligned}$$

Recordemos que por las hipótesis del modelo de Gauss-Markov, el vector de residuos es ortogonal sobre todos los regresores de su modelo. Esto significa que para el modelo restringido (donde hemos eliminado X_2) $X_1'\hat{\varepsilon}_R = 0$ pero $X_2'\hat{\varepsilon}_R \neq 0$ porque X_2 no forma parte del modelo y por tanto no se tiene que cumplir la hipótesis de ortogonalidad; para el modelo completo $X_1'\hat{\varepsilon} = 0$ y $X_2'\hat{\varepsilon} = 0$. Aplicamos esta propiedad para obtener $X_1'\hat{\varepsilon} = 0$ y por propiedad de matrices sabemos que $(X_1'X_1)^{-1}X_1'X_1 = I$. Queda entonces que

$$\hat{\beta}_R = \hat{\beta}_1 + \overbrace{(X_1'X_1)^{-1}X_1'X_2}^Q \hat{\beta}_2.$$

De esta ecuación deducimos que la diferencia entre $\hat{\beta}_1$ y $\hat{\beta}_R$ viene dada por el valor de Q y de $\hat{\beta}_2$ del modelo completo. Si alguno de estos dos elementos es 0, entonces se daría que $\hat{\beta}_R = \hat{\beta}_1$. Esto ocurre únicamente si las variables explicativas X_2 no explican al modelo, es decir, $\hat{\beta}_2 = 0$, o si resulta que X_1 y X_2 son ortogonales, lo que implicaría que $X_1'X_2 = 0$.

Si se diese el caso que $\hat{\beta}_R = \hat{\beta}_1$ entonces daría lo mismo hacer la regresión en el modelo completo o en el restringido, aunque en la realidad lo común es que $\hat{\beta}_R \neq \hat{\beta}_1$.

❖ Comparación entre: $\hat{\varepsilon}'\hat{\varepsilon}$ y $\hat{\varepsilon}'_R\hat{\varepsilon}_R$.

A continuación, se estudiará la diferencia entre los estimadores de las perturbaciones del modelo completo $\hat{\varepsilon}$ y del modelo restringido $\hat{\varepsilon}_R$. El modelo completo contiene más variables explicativas que el modelo restringido, por lo que es de esperar que proporcione unos errores menores o iguales al modelo restringido, es decir,

$$\hat{\varepsilon}'\hat{\varepsilon} \leq \hat{\varepsilon}'_R\hat{\varepsilon}_R.$$

Para demostrar esta afirmación comenzamos con recordar que $\hat{\varepsilon}_R = y - X_1\hat{\beta}_R$. Debemos realizar las siguientes transformaciones. Como $\hat{\beta}_R = (X_1'X_1)^{-1}X_1'y$, sustituimos y obtenemos

$$\hat{\varepsilon}_R = y - X_1\hat{\beta}_R = y - X_1 \overbrace{((X_1'X_1)^{-1}X_1'y)}^{\hat{\beta}_R}.$$

A continuación, sacamos factor común de y y obtenemos

$$\hat{\varepsilon}_R = (I - X_1(X_1'X_1)^{-1}X_1')y = M_1y,$$

donde $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$. Seguidamente sustituimos y por su expresión del modelo completo obteniendo así

$$\hat{\varepsilon}_R = M_1y = M_1(X_1\hat{\beta}_1 + X_2\hat{\beta}_2 + \hat{\varepsilon}) = \overbrace{M_1X_1\hat{\beta}_1}^0 + M_1X_2\hat{\beta}_2 + \overbrace{M_1\hat{\varepsilon}}^{\hat{\varepsilon}}.$$

Como $M_1X_1 = [(I - X_1(X_1'X_1)^{-1}X_1')X_1 = X_1 - X_1(X_1'X_1)^{-1}X_1'X_1 = X_1 - X_1] = 0$, $M_1\hat{\varepsilon} = (I - X_1(X_1'X_1)^{-1}X_1')\hat{\varepsilon} = \hat{\varepsilon} - X_1(X_1'X_1)^{-1}X_1'\hat{\varepsilon} = \hat{\varepsilon}$ porque $X_1'\hat{\varepsilon} = 0$, entonces

$$\hat{\varepsilon}_R = M_1X_2\hat{\beta}_2 + \hat{\varepsilon}.$$

Por tanto, la diferencia entre $\hat{\varepsilon}_R$ y $\hat{\varepsilon}$ recae en $M_1X_2\hat{\beta}_2$. Podemos observar que $\hat{\varepsilon}_R = \hat{\varepsilon}$ sólo si $M_1X_2\hat{\beta}_2 = 0$, que se cumple cuando las variables explicativas contenidas en X_2 no explican la variable explicada, es decir, cuando $\hat{\beta}_2 = 0$.

Cuando realizamos la comparación con los residuos al cuadrado para omitir los valores negativos, obtenemos que

$$\hat{\varepsilon}_R'\hat{\varepsilon}_R = \hat{\beta}_2'X_2'M_1X_2\hat{\beta}_2 + \hat{\varepsilon}'\hat{\varepsilon}.$$

Como M_1 es una matriz semidefinida positiva, se tiene que

$$\hat{\beta}_2'X_2'M_1X_2\hat{\beta}_2 = (X_2\hat{\beta}_2)'M_1(X_2\hat{\beta}_2) \geq 0,$$

es decir, que la diferencia entre $\hat{\varepsilon}_R'\hat{\varepsilon}_R$ y $\hat{\varepsilon}'\hat{\varepsilon}$ será siempre cero o positiva, por lo que queda demostrado que

$$\hat{\varepsilon}'\hat{\varepsilon} \leq \hat{\varepsilon}_R'\hat{\varepsilon}_R.$$

❖ **Omitir variables o mantenerlas.**

Normalmente pueden existir una inmensa cantidad de posibles variables explicativas para una variable explicada de carácter económico, por lo que la pregunta consiste en saber qué variables explicativas incluir en el modelo de estudio. Lo lógico es que se incluyan sólo aquellas que tengan un claro impacto sobre la variable explicada y eliminar aquellas que sean menos importantes.

Para analizar el impacto de eliminar unas variables u otras vamos a apoyarnos en la premisa de que conocemos el modelo verdadero y completo definido como

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

El análisis consistirá en determinar qué ocurre cuando sólo consideramos las variables explicativas de X_1 omitiendo así X_2 , aunque esta última pertenezca al modelo verdadero. Como se ha visto antes para este caso una de las expresiones del estimador de β_1 del modelo restringido, $\hat{\beta}_R$, es

$$\hat{\beta}_R = \hat{\beta}_1 + \overbrace{(X_1'X_1)^{-1}X_1'X_2}^Q \hat{\beta}_2,$$

donde se decía que la diferencia entre $\hat{\beta}_R$ y $\hat{\beta}_1$ venía dada por Q y $\hat{\beta}_2$. Sobre esta expresión vamos a aplicar las esperanzas, obteniendo así

$$E[\hat{\beta}_R] = E[\hat{\beta}_1] + QE[\hat{\beta}_2].$$

Como $\hat{\beta}_1$ y $\hat{\beta}_2$ son estimadores insesgados del modelo completo verdadero, se tiene que

$$E[\hat{\beta}_R] = \beta_1 + Q\beta_2.$$

De esta expresión podemos deducir que $\hat{\beta}_R$ no es un estimador insesgado del modelo restringido y que su sesgo de variables omitidas viene dado por Q . Sólo sería un estimador insesgado si el modelo restringido fuese el verdadero, y por tanto $\beta_2 = 0$.

3.5. Inferencia

En este apartado se tratará de determinar si o más variables explicativas deben mantenerse en el modelo o eliminarse. Para ello se proponen dos hipótesis con la intención de concluir se debe aceptar o rechazar la primera hipótesis, denominada hipótesis nula (H_0), por medio de los test T de Student y F de Snedecor.

❖ **Test T de Student.**

El test T de Student estudia la significación de los coeficientes individuales (*estimadores* β_j , $j = 2, \dots, k$) de las variables explicativas. Las hipótesis planteadas son las siguientes:

$$H0: \beta_j = 0, \quad H1: \beta_j \neq 0,$$

es decir, la hipótesis a contrastar es si el coeficiente de determinada variable es igual a cero y por tanto es insignificante, pudiéndose eliminar del modelo. En caso de no poder rechazar la hipótesis nula, se acepta la alternativa.

Para realizar el contraste se calcula un estadístico denominado t -valor para cada variable explicativa que se desee contrastar bajo las anteriores hipótesis. Para esto se emplea la siguiente ecuación:

$$t_j = \frac{\hat{\beta}_j}{s_j}.$$

Como podemos observar, para calcular t_j se requiere del estimador insesgado de β_j y de la varianza de la perturbación σ^2 , $\hat{\beta}_j$ y s_j , obtenidos anteriormente. Recordemos que $s_j = s\sqrt{a_{jj}}$, donde a_{jj} es el j -ésimo elemento de la diagonal de la matriz $(X'X)^{-1}$, por lo que podemos escribir también la expresión de t_j como

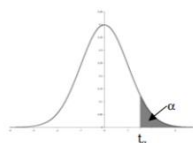
$$t_j = \frac{\hat{\beta}_j}{s\sqrt{a_{jj}}}.$$

A continuación, se debe tomar la tabla de la distribución T de Student y buscar el valor con el que comparar t_j y determinar si se debe rechazar o no la hipótesis nula. Para ello, primero se selecciona un nivel de significación α de forma arbitraria, normalmente 0.05.

Este es un fragmento de la tabla T de Student:

PERCENTILES DE UNA DISTRIBUCIÓN t DE STUDENT

La tabla proporciona los percentiles de una distribución t de Student con v grados de libertad y probabilidad α .



Nota: El valor de cada celda sólo considera una cola de la distribución.

$\alpha=$	0,100	0,050	0,030	0,025	0,020	0,015	0,010	0,005	0,0010	0,0005
$v=1$	3,078	6,314	10,579	12,706	15,894	21,205	31,821	63,656	318,289	636,578
2	1,886	2,920	3,896	4,303	4,849	5,643	6,965	9,925	22,328	31,600
3	1,638	2,353	2,951	3,182	3,482	3,896	4,541	5,841	10,214	12,924
4	1,533	2,132	2,601	2,776	2,999	3,298	3,747	4,604	7,173	8,610
5	1,476	2,015	2,422	2,571	2,757	3,003	3,365	4,032	5,894	6,869
6	1,440	1,943	2,313	2,447	2,612	2,829	3,143	3,707	5,208	5,959
7	1,415	1,895	2,241	2,365	2,517	2,715	2,998	3,499	4,785	5,408
8	1,397	1,860	2,189	2,306	2,449	2,634	2,896	3,355	4,501	5,041
9	1,383	1,833	2,150	2,262	2,398	2,574	2,821	3,250	4,297	4,781
10	1,372	1,812	2,120	2,228	2,359	2,527	2,764	3,169	4,144	4,587
11	1,363	1,796	2,096	2,201	2,328	2,491	2,718	3,106	4,025	4,437
12	1,356	1,782	2,076	2,179	2,303	2,461	2,681	3,055	3,930	4,318
13	1,350	1,771	2,060	2,160	2,282	2,436	2,650	3,012	3,852	4,221
14	1,345	1,761	2,046	2,145	2,264	2,415	2,624	2,977	3,787	4,140
15	1,341	1,753	2,034	2,131	2,249	2,397	2,602	2,947	3,733	4,073

Las filas representan los grados de libertad. Para obtenerlo basta con calcularlo por medio de la expresión

$$gdl = n - k.$$

El valor que buscamos de la tabla, $t(n - k)$, correspondiente al asociado a la fila $n - k$ y la columna correspondiente a la mitad del nivel de significación, comúnmente 0.025.

Si $-t(n - k) < t_j < t(n - k)$ entonces no se puede rechazar H_0 porque t_j cae en la región de no rechazo (*color blanco en la gráfica de la tabla*). Si está fuera de dicho intervalo (*color gris*), entonces se rechaza H_0 y se toma por válida H_1 .

❖ Test F de Snedecor.

El test F de Snedecor es una generalización del test T de Student donde, en lugar de estudiar la significación de una única variable explicativa, se estudia la significación de un varias conjuntamente mediante la comparación entre dos modelos: uno completo (*entendiendo por completo como el modelo con todas las variables que el restringido elimina*) y otro restringido, poniendo en contraste la significación del conjunto de variables β_2 proveniente del modelo

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

Las hipótesis son las siguientes:

$$H_0: \beta_2 = 0, \quad H_1: \beta_2 \neq 0,$$

mientras que el estadístico que se calcula en este test se denomina F -valor. Si se rechaza la hipótesis nula, entonces es más conveniente mantener el modelo completo. Como se puede observar, no se están especificando significaciones de variables concretas, por lo que la hipótesis H_1 viene a decir que al menos una de las variables que contiene es significativa.

El F -valor puede obtenerse mediante la expresión

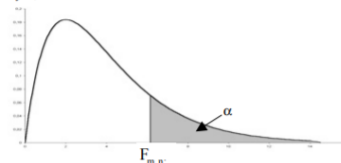
$$F = \frac{n - k}{g} \cdot \frac{R^2 - R_R^2}{1 - R^2},$$

donde g representa el número de variables explicativas que el modelo restringido elimina del modelo completo.

A continuación, se debe tomar la tabla de la distribución T de Student y buscar el valor con el que comparar F junto con un nivel de significación α determinado. Un fragmento de esta tabla es el siguiente:

PERCENTILES DE UNA DISTRIBUCIÓN F DE SNEDECOR CON m Y n GRADOS DE LIBERTAD

La Tabla proporciona los percentiles de una distribución F de Senedecor con m grados de libertad en el numerador y n grados de libertad en el denominador. Para cada par de valores m y n aparecen cuatro valores que corresponden a las probabilidades $\alpha=0,05$; 0,025; 0,01 y 0,001. Para distinguir el valor correspondiente a $\alpha=0,025$ aparece en negrita.



m=	1	2	3	4	5	6	7	8	9	10	12,00	24	∞
n=1	161,45 647,79 4.052,18 405.311,58	199,50 799,48 4.999,34 499.725,34	215,71 864,15 5.403,53 540.256,50	224,58 899,60 5.624,26 562.667,85	230,16 921,83 5.763,96 576.496,12	233,99 937,11 5.858,95 586.032,87	236,77 948,20 5.928,33 593.185,42	238,88 956,64 5.980,95 597.953,80	240,54 963,28 6.022,40 602.245,33	241,88 968,63 6.055,93 605.583,19	243,90 976,72 6.106,68 610.351,56	249,05 997,27 6.234,27 623.703,00	254,19 1.017,76 6.362,80 636.100,77
2	18,51 38,51 98,50 998,38	19,00 39,00 99,00 998,84	19,16 39,17 99,16 999,31	19,25 39,25 99,25 999,31	19,30 39,30 99,30 999,31	19,33 39,33 99,33 999,31	19,35 39,36 99,36 999,31	19,37 39,37 99,38 999,31	19,38 39,39 99,39 999,31	19,40 39,40 99,40 999,31	19,41 39,41 99,42 999,31	19,45 39,46 99,46 999,31	19,49 39,50 99,50 999,31
3	10,128 17,443 34,116 167,056	9,552 16,044 30,816 148,488	9,277 15,439 29,457 141,095	9,117 15,101 28,710 137,079	9,013 14,885 28,237 134,576	8,941 14,735 27,911 132,830	8,887 14,624 27,671 131,608	8,845 14,540 27,489 130,618	8,812 14,473 27,345 129,861	8,785 14,419 27,228 129,221	8,745 14,337 27,052 128,319	8,638 14,124 26,597 126,932	8,529 13,908 26,137 123,517

En este caso, las filas corresponden al valor g y las columnas a $n - k$.

Si $0 < F < F(g, n - k)$, entonces F cae en la región de no rechazo (*color blanco de la gráfica*) y no se puede rechazar H_0 . Sin embargo, si $F > F(g, n - k)$, F cae en la región de rechazo (*color gris*), y por tanto se rechaza H_0 concluyendo que el modelo restringido es el más significativo.

En ocasiones interesa comparar el modelo completo con el modelo constante, es decir, donde se eliminan todas las variables explicativas dejando únicamente el intercepto. En este caso, el test F se denomina test de significación global, y $g = k - 1$.

El test F es equivalente al test T cuando el primero compara un modelo completo con otro restringido donde únicamente se elimina una variable, siendo $F_{1,n-k} = t_{n-k}^2$ ($g = 1$ porque sólo se elimina una variable) y los p -valores de ambos estadísticos exactamente iguales. Sin embargo, en estos casos es preferible abordar el contraste de hipótesis mediante el test T, puesto que el cálculo del t -valor es más sencillo.

Bibliografía

- [1] C. Heij, P. de Boer, P. H. Franses, T. Kloek y H. K. van Dijk, *Econometric Methods with Applications in Business and Economics*, New York: Oxford University Press, 2004.